Video Demo: An Egocentric Vision based Assistive Co-robot

Jingzhe Zhang[#], Lishuo Zhuang[§], Yang Wang[§], Yameng Zhou[#], Yan Meng[#], Gang Hua[§]

[#]Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ 07030
{jzhang23, yzhou5, yan.meng}@stevens.edu

[§]Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ 07030
{ywang7,lzhuang,ghua}@stevens.edu

Abstract

We present the video demo of the prototype of an egocentric vision based assistive co-robot system. In this co-robot system, the user is wearing a pair of glasses with a forward looking camera, and is actively engaged in the control loop of the robot in navigational tasks. The egocentric vision glasses serve for two purposes. First, it serves as a source of visual input to request the robot to find a certain object in the environment. Second, the motion patterns computed from the egocentric video associated with a specific set of head movements are exploited to guide the robot to find the object. These are especially helpful for quadriplegic individuals who do not have the needed hand functionality for control with other modalities (e.g., joystick). In our co-robot system, when the robot does not fulfill the object finding task in a pre-specified time window, it would actively solicit user controls for guidance. Then the users can use the egocentric vision based gesture interface to orient the robot towards the direction of the object. After that the robot will automatically navigate towards the object until it finds it. Our experiments validated the efficacy of the closed-loop design to engage the human in the loop.

1. Introduction

Robotic technology has the potential to offer considerable help for those disabled individuals. Rehabilitation and assistive robotic systems, like Care-O-bot II [1], usually have a hand operating interface, e.g., touch screen or joystick. However, for quadriplegic individuals who lost their hand functionality, it is very important to explore other modalities to design a human robot interface that they actually operate to control such assistive robotic systems.

We present a prototype co-robot system that is equipped with a pair of egocentric video glasses, which captures the visual scene from the first person view (FPV). The egocentric camera built on the glasses captures the head movement gesture of the user, which are exploited to guide the robot finding the target object, which provides a hand-free user interface for those people who are unable to operate the co-robot by their hands. Simplifying operations of those severely disabled, only several natural head gestures such as "nodding", "shaking head", and "turning head" around have been considered to be included in the gesture lexicon to trigger interactive commands in our human-robot interaction system. The intended application of this co-robot system is to help a person with severe disability for grabbing a target object. Since grabbing by itself is a very intricate task, in this paper, our prototype co-robot system focused on the task of object finding, with the user in the loop.

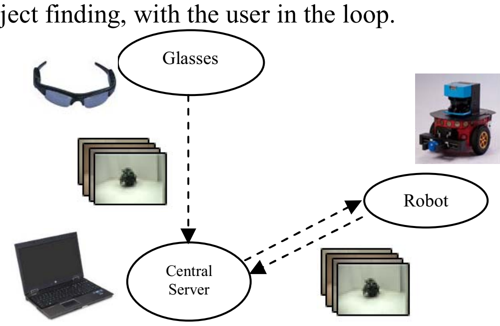

Fig. 1. The main components of the overall assistant co-robot system.

In this co-robot system, three main components are engaged: a pair of egocentric vision glasses, a central server (CS), and a mobile Pioneer P3-DX robot, as shown in Fig. 1. The user looks at the target object and nod his/her head, triggering to take a picture of the target object and simultaneously send a command to the co-robot through a central server to find this object. If the robot cannot find the object within a pre-specified time period, it will ask for the user's help through speech, *i.e.*, by saying "I need help". Then the user would use the head motion gesture to guide the robot towards the direction of the object and then using the nod gesture issuing a "continuing search in that direction" command. The robot then automatically navigates towards the object, avoiding any obstacles during the navigating process.

For the purpose of mobility, the three components are all connected via a wireless network. A finite-state machine of the system is shown in Fig. 2, and 5 unique states are designed to delimit all possible conditions in the co-robot system. The switches among different states are only triggered either by the CS policy or by the head movements from the egocentric vision glasses. Specifically, "turning head around as a circle" accesses object recognition mode from the idle mode, shaking head quits from the running robot search or the target approach mode to the idle, turning head horizontally guides the same turning motions of robot to the target direction in the FPV glasses control mode, and nodding head confirms the robot moving forward to the direction, to list a few.

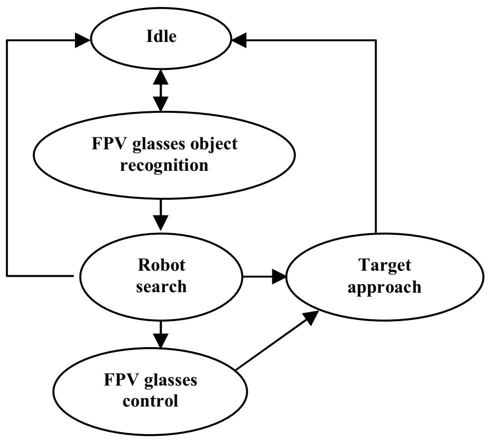

Fig. 2: System finite state machine

2. System Components

The Egocentric Vision Glasses: The egocentric vision glasses are a pair of sunglasses with a forward looking camera in the middle, with a view angle of 60 degrees. The user wearing the egocentric vision glasses looks around, making the view of the glasses camera be roughly equal to the field-of-view of the user. The co-robot system understands the intentions of the user according to the videos taken from this egocentric camera, and makes proper decisions accordingly. The WIFI module of the glasses enables the communication between the video glasses and the CS via 2.4 GHz wireless channel on the fly. We fetch the video with the resolution of 640×480.

Motion-pattern Estimation: The motion-pattern estimation module composes a 56-dimension motion histogram to estimate the motion pattern of the egocentric vision glasses. Every shifting distance between two adjacent image frames is evaluated by Full-search Block Matching algorithm after down-sampling. These Shifting distances are stored in a circular buffer sequentially. All motion vectors (x, y) in the circular buffer are categorized into 8 directions. Both 16-demision and 32-demision histograms are generated by producing binary partition twice on the circular buffer and the features are concatenated to better reserve the sequential information of frames. Therefore, the combination of all the three histograms, a 56-dimension histogram, contains both global and local information. A support vector machine (SVM) is trained on top of the 56-dimension vectors from a set of labeled training data to recognize different motion gestures.

Object Recognition: We employ the vocabulary tree (VT) based recognition algorithm proposed by Nister [3] for fast indexing and retrieval of relevant object images and a Homography-based post-verification is conducted to improve the recognition accuracy and identify the local space information of objects in a specific image.

Two databases are independently established to address variance of the egocentric vision glasses and the robot camera. Top 4 candidates from the VT are fed into Homography-based post verification scheme to re-evaluate and finalize the decision. We use multi-threading based on Open MP (Open Multiprocessing), which significantly reduces the execution time of the post verification processes. Two constraints are checked, in order to verify the accuracy of the Homography, the scale and the similarity of distribution of object key-points.

Robot Localization, Mapping and Path-Planning: We use a Pioneer 3-DX mobile robot in our proof-of-concept assistive co-robot system. The mission of this robotic system is to control the robot to always move on the shortest path and hence in the most efficient way. We basically rely on the Mobile Robots' Advanced Robot Interface for Applications (Aria) and Advanced Robot Navigation and Localization (ARNL), to accomplish robot localization, map building, and path planning. We also designed an effective approaching algorithm to decide an appropriate route to the target object based on these libraries after the object is recognized.

3. Conclusion

Our experiments in an object finding task clearly demonstrated the efficacy of the human-in-the-loop strategy, which is 3 times and 6 times better than tour search, searching from predefined spots in order, and random search, depending on whether there are obstacles presented or not. We refer interested readers to [3], for more technical details of this egocentric vision based co-robot system.

References

[1] Graf, B., Hans, M., & Schraft, R. D. Care-O-bot II—Development of a Next Generation Robotic Home Assistant. *Autonomous Robots*, pp. 193–205, 2004.

[2] David Nist´er and Henrik Stew´enius. Scalable Recognition with a Vocabulary Tree. In CVPR, 2006.

[3] Jingzhe Zhang, Lishuo Zhuang, Yang Wang, Yameng Zhou, Yan Meng, Gang Hua, "An Egocentric Vision based Assistive Co-robot", In Proc. *International Conf. on Rehabilitation Robotics*, 2013.